

Freeway truck travel time prediction for freight planning using truck probe GPS data

Zun Wang¹

University of Washington, United States.

Anne V. Goodchild²

Department of Civil and Environmental Engineering, University of Washington, United States.

Edward McCormack³

Department of Civil and Environmental Engineering, University of Washington, United States.

Predicting truck (heavy vehicle) travel time is a principal component of freight project prioritization and planning. However, most existing travel time prediction models are designed for passenger vehicles and fail to make truck specific forecasts or use truck specific data. Little is known about the impact of this limitation, or how truck travel time prediction could be improved in response to freight investments with an improved methodology. In light of this, this paper proposes a pragmatic multi-regime speed-density relationship based approach to predict freeway truck travel time using empirical truck probe GPS data (which is increasingly available in North American and Europe) and loop detector data. Traffic regimes are segmented using a cluster analysis approach. Two case studies are presented to illustrate the approach. The travel time estimates are compared with the Bureau of Public Roads (BPR) model and the Akçelik model outputs. It is found that the proposed method is able to estimate more accurate travel times than traditional methods. The predicted travel time can support freight prioritization and planning.

Keywords: *truck travel time prediction, truck GPS data, multi-regime speed-density, cluster analysis.*

1. Introduction

In both Europe and North American, predicting truck (heavy vehicle) travel time is a principle component of freight planning. For instance, most freight prioritization tools count travel time reduction as one of the key project benefits associated with a freight investment. Travel time changes are also an input into other calculations, for example, vehicle operating cost. Historically, vehicle performance functions have been used to predict travel time in planning tools, e.g. project prioritization tools and travel demand models. However, most vehicle performance functions are designed to represent passenger travel and do not consider truck performance separately. As a result, in these planning tools, trucks performance is either treated the same as passenger travel or approximated by simply applying an adjustment factor to passenger travel. For instance, the Puget Sound Regional Council's (in Washington State, USA) travel demand model converts truck volume to passenger car equivalents for trip assignment and applies an additional 25% factor on travel time of trucks traveling on freeways during model calibration (Cambridge Systematics,

¹ A: 121E More Hall, University of Washington, Seattle, United States T: +01 352-665-1286 E: zunwang@uw.edu

² A: 121E More Hall, Box 352700, University of Washington, Seattle, United States T: +01 206-543-3747 F: +01 206 543-1543 E: annegood@u.washington.edu

³ A: 121C, More Hall, University of Washington, Seattle, United States, T: +01 206-543-3348 E: edm@uw.edu

2007). Similarly, the Atlanta Regional Commission (in Georgia State, USA) model assigns trucks to the network with a time-penalty value relative to passenger travel (Atlanta Regional Commission 2011). Although there are a considerable number of truck specific models, they are designed for modelling truck demand generation and distribution (Cambridge Systematics and Jack Faucett Associates, 2001) and truck specific performance functions for predicting travel time are not found in the literature. One reason for this is the deficiency of truck specific data (McCormack and Hallenbeck 2006). What's more, many passenger trip travel time prediction models that have been used for truck related study rely on out-of-date data or limited samples. For instance, the travel time prediction model developed by the U.S. Bureau of Public Roads, called the BPR function, was proposed based on data collected on uncongested highways, and therefore is not able to capture the travel time under congested conditions (U.S. Bureau of Public Roads, 1964).

Fortunately, the needs for quantitative freight performance measures and planning have been recognized. An increasing number of trucks in both North America and Europe are equipped with GPS devices (McCormack and Aakre 2014). One market research firm forecasts that by 2017 21% (Europe) and 22% (North America) of all non-privately owned commercial vehicles will have GPS fleet management systems (Berg Insight 2014). Several studies have investigated how to collect, process and apply truck GPS data in North America, Australia and European countries (McCormack and Hallenbeck 2006, Greaves and Figliozzi 2008, and Pluvient et al. 2012). While most truck GPS data based research focus on evaluating current performance, no known study predicts truck travel time using truck GPS data. Thus the objective of this research is to propose a transferable framework to predict truck travel time that can be applied to different locations based upon available truck GPS data. The models generated by the proposed framework can support freight planning, e.g. estimating future truck travel time associated with freight investment or predicting truck travel time in travel demand model.

2. Literature Review

2.1 State of the practice of truck probe GPS data applications

The usage of truck GPS data for freight related studies has gained increasing attention given the growing market penetration of GPS technology. A number of transportation agencies have or are planning to use truck GPS data to evaluate roadway performance. For example in Norway, data from GPS devices will be used to monitor important freight corridors (McCormack and Aakre 2014). In the USA, the Federal Highway Administration (FHWA) collaborated with the American Transportation Research Institute (ATRI) to investigate how data gathering from GPS devices installed in trucks can be used to measure mobility and reliability along U.S. interstate highways (ATRI and FHWA 2005). A series of studies have been conducted since 2002, including measuring average truck speed, travel time, and travel time reliability (using the buffer index). By 2012, 250 freight-significant highway locations in the U.S. were monitored and ranked based on congestion using truck GPS data. In June 2013, the FHWA released the National Performance Management Research Data Set, a 5-minute aggregated truck GPS speed data set covering the entire US national highway system for truck performance measure (FHWA 2013). As truck GPS data is increasingly available to transportation agencies and researchers, there are studies evaluating truck mobility. Figliozzi et al. (2011) utilized the ATRI truck GPS data to examine the travel time and travel time reliability on Interstate-5 corridor in Oregon State, USA. Using statistical techniques. Liao (2014) studied truck mobility, delay and reliability along 38 critical freight-significant corridors in Twin-City, USA to identify truck bottlenecks using truck GPS data. Ma et al. (2011) implemented a truck trip identification algorithm based on truck GPS data collected in the City of Seattle, USA. They also developed an online platform to measure and report truck trip performance including speed, trip distance, travel time, and travel time reliability. Zhao et al. (2011) employed the same dataset to measure truck travel time on

freeways. Wang et al (2015) employed truck GPS data to quantify truck travel time reliability using different models and provided recommendations on the appropriate models under different conditions.

In addition, truck GPS data is used to understand truck trip characteristics. Pluvinet et al. (2012) proposed a GPS survey methodology to collect truck movement data in Spain and France. The GPS data was collected by mobile phone devices. The authors analyzed the trip characters, including number of delivery stops, duration of a delivery stop, length of route, and duration of a route. Truck GPS data also had been used to understand and evaluate truck behaviors, e.g. the studying by Wang and Goodchild (2014) quantify the impacts of tolling on truck speed and routing choice. Despite that there are studies using truck GPS data to study truck mobility and behaviors, none of them provide insight into how truck GPS data can be used to predict future truck travel time.

2.2 Prediction of travel time

There is considerable research being done on predicting travel time (Lin et al 2005). These approaches can be classified into two categories based on their applications: short-term (real time) travel time prediction for traffic operation purposes and long-term travel time estimation for transportation planning purposes. A great deal of recent research has been targeted at developing short-term travel time prediction models using statistical techniques and mathematical modelling approaches, including time series (D'Angelo 1999), Kalman filtering (Chien 2003), artificial neural networks (Van Lint 2005), and Markov chain (Yeon 2008). Most of these approaches require current traffic conditions and historical observations, as well as considerable computing resources to develop predictions for real-time traffic operations. The objective of this paper is to propose an approach that can support long-term freight project prioritization and planning, not real-time operations, and therefore, the literature review emphasizes travel time prediction over a longer time horizon.

One of the most straightforward methodologies for longer time travel time estimation is the use of speed and volume-capacity ratio (V/C) relationship. It has been applied extensively in various project benefit-cost analysis tools (McFarland 1993, Dowling Associates 2000). The speed is predetermined and changes in response to various V/C, facility type and speed limit. This engineering relationship is simple but not always accurate. In addition, it does not capture any network effects when additional traffic is attracted to the improved segments from other roads. Equilibrium traffic assignment methods address this issue, by assigning traffic to the network based on the predefined cost functions. The entire system reaches an equilibrium status assuming all vehicles travel along the minimum cost path. For instance, the Freight Analysis Framework version 3 (FAF3) freight traffic analysis developed by Battelle (2011) uses this method to assign freight traffic flow to the national highway network. The FAF3 employs the BPR function (U.S. Bureau of Public Roads 1964) as the cost function for the stochastic user equilibrium traffic assignment procedure, as shown in Equation 1.

$$TT_{BPR} = TT_{ff} \times [1 + a(x)^b] \quad (1)$$

where TT_{BPR} = segment traversal time estimated using the BPR function,

TT_{ff} = segment vehicle travel time at free flow speed,

x = volume-capacity ratio,

a and b are determined by facility type, free-flow speed and speed at capacity.

According to the Highway Capacity Manual (HCM 2000), the freeway free flow speed is calculated based on the information of number of lanes, lane width, shoulder width, and interchange density. Segment capacity is defined as number of vehicles during one hour under free-flow condition, and determined by facility type and free-flow speed.

The parameter α of the BPR function influences the ratio of free-flow speed to the speed at capacity. The parameter β determines how sensitive the speed change is when v/c is close to 1.0 (Dowling et al 1998). Given the characteristics of the two case studies of this paper, α and β are assigned to 0.15 and 4 respectively.

The BPR function assumes travel time has a linear relationship with volume-capacity ratio. The model was developed by fitting data collected on uncongested freeways, and does not capture the travel time under congestion condition. To overcome the inaccurate prediction of oversaturated condition, Akçelik developed a time-dependent travel time prediction function based on the steady-state delay equation for a single channel queuing system, and this model was recommended by the HCM 2000 for predicting vehicle travel time for planning purposes (Akçelik et al 1991, HCM 2000). The model is shown in Equation 2.

$$TT_{Akçelik} = TT_{ff} + 0.25T[(x-1) + \sqrt{(x-1)^2 + \frac{16JL^2}{T^2}}] \quad (2)$$

where $TT_{Akçelik}$ = segment traversal time predicted using the Akçelik function,

T = expected duration of demand (typically 1 hour),

L = segment length (mile)

J = calibration parameters determined by facility type, signal per mile, free-flow speed and speed at capacity (exhibit 30-4, HCM 2000)

Although the above two travel time prediction equations are extensively employed in travel demand models to estimate vehicle speed and travel time in response to various traffic volumes, neither of them is a truck specific model. As a result, there exist considerable deviations between the truck travel estimates and actual truck travel times. To solve this issue and predict reasonably accurate truck travel time for freight planning, this paper proposes an approach to forecast truck travel time based on empirical truck GPS observations.

The above literature review reveals that (1) despite the existence of studies using truck GPS data to study truck mobility and behaviors, none provide insight into how truck GPS data can be used to predict truck travel time, (2) although there exist travel time prediction equations that are extensively employed in travel demand models to estimate vehicle speed and travel time in response to various traffic volumes, none of them are truck specific. To bridge this gap, this paper proposes a pragmatic approach to estimate truck travel time in response to traffic changes using truck GPS data and loop data. The logic of this approach is based on multi-regime relationships between truck speed and segment density. Cluster analysis was employed to segment traffic regimes, and truck travel time is estimated in response to segment density changes. These predicted travel times can be used to estimate travel time changes associated with freight investments or other planning practices.

3. Methodology

This section introduces the proposed freeway truck travel time prediction approach. The proposed approach predicts truck travel time based on the relationships between truck speed and density, which were retrieved from truck GPS data and dual-loop detectors respectively. The k-means cluster analysis algorithm was selected to partition data into homogeneous groups based on the characteristics of different traffic regimes.

The travel time prediction approach consists of 4 major steps:

1. Classify clusters based on the characteristics of truck speed and segment traffic volume using k-means algorithm,
2. Fit speed-density relationships,

3. Estimate freeway truck travel time, and
4. Evaluate estimation accuracy.

3.1 Identify clusters

Existing traffic flow studies have observed that traffic data shows two clear phases: free-flow and congested phases. In the free-flow phase, vehicles move at their desired speed and there is little influence/interaction between vehicles. In the congested phase, the traffic volume on the segment approaches capacity, and vehicles speed declines. Recent studies have also identified a transitional phase, called the intermediate phase (Kerner 1996). In the intermediate phase, vehicles experience stop-and-go driving conditions and are forced to drive as part of the overall traffic. Both two-regime and three-regime traffic models have been proposed in the literature. The first two-regime traffic flow model was proposed by Edie (1961), in which, the free-flow regime was fitted using the Underwood model and the congestion-flow regime was represented by Greenberg model, as shown in Equation (3).

$$u = \begin{cases} 54.9 \exp(-k / 163.9) & \text{for } k \leq 50 \\ 26.8 \ln(162.5 / k) & \text{for } k \geq 50 \end{cases} \quad (3)$$

where u = vehicle speed (mph)

k = traffic density (vehicles per lane per mile)

Drake et al. (1967) developed a three-regime traffic model based on the Greenshields-type linear model for all three regimes, as given in Equation (4).

$$u = \begin{cases} 50 - 0.098k & \text{for } k \leq 40 \\ 81.4 - 0.913k & \text{for } 40 \leq k \leq 65 \\ 40 - 0.265k & \text{for } k \geq 65 \end{cases} \quad (4)$$

While these multi-regime models substantially improve the capability to capture different traffic characteristics under various traffic conditions, one of the major challenges of proposing such models is to determine the breakpoints between regimes (Sun and Zhou 2005). In the literature, most density breakpoints were determined by the researchers' engineering experience, which is subjective and biased by the judgment of model developers. Sun and Zhou (2005) employed the Cluster Analysis method to determine the breakpoints automatically given the fact that data belongs to the same cluster share similar features and data with different features belong to different groups. This paper also employs a cluster analysis method to determine the breakpoints. Cluster analysis is a methodology to classify samples into a number of groups using a quantitative measure of association. The k-means algorithm is chosen in this study to identify traffic clusters. This algorithm is a centroid-based clustering algorithm, which aims to find the k cluster centers and assign the data to the nearest cluster center whose mean yields the least within-cluster sum of squares (Hartigan 1975). The k-means algorithm requires that the number of clusters is predetermined by modelers. Other cluster methods which the number of clusters is not predetermined (e.g. hierarchical clustering) were not selected. This is because the objective of this study is to develop an applicable approach to classify data belonging to the same traffic regime and to develop speed-density relationships, rather than calculating the optimal number of clusters that may not be applicable to establish speed-density relationships. The cluster analysis was accomplished using the R software package "cluster" (R Software 2014). Clusters are identified by minimizing the distance between observations and centroid of each cluster.

3.2 Fit truck speed-density relationships

For each cluster, the corresponding speed density relationship is fitted by minimizing squared errors. According to empirical observations, the speed-density relationships usually follow three relationships. The linear relationship originated from the Greenshields' Model (Greenshields

1935); the logarithmic relationship originated from the Greenberger Model (Greenberger 1957), and the exponential relationships originated from the Underwood Model (Underwood 1961). The three original models are presented in Equation 5 to 7. The appropriate format to fit the data is determined based on the adjusted R-squared values using the R software. The one with the greatest R-squared value is chosen to represent the speed-density relationship of the empirical observations. The major advantage of these three models is the mathematical simplicity, while the primary drawback is none of these three models is able to draw a whole picture of speed-density relationship. More specifically, the Greenshields' model assumes speed and density follows a linear relationship, but it is hard to find such a relationship in empirical data. The Greenberger model is not able to predict speed at low density where the speed tends to infinity when density tends to infinity. For the Underwood model, it fails to predict speed at high density situation where the speed is equal to zero when density reaches infinity (Wang 2009).

$$u = u_f (1 - k / k_j) \quad (5)$$

$$u = u_m \ln(k_j / k) \quad (6)$$

$$u = u_f \exp(-k / k_0) \quad (7)$$

Where u_f = free-flow speed (mph)

k_j = jam density (vehicles per lane per mile)

u_m = optimal speed when flow reaches the maximum value (mph)

k_0 = optimal density (vehicles per lane per mile)

3.3 Estimate truck travel time

Truck travel time is estimated by dividing segment distance by speed predicted on the speed-density relationships. It is assumed that trucks travel at a constant speed along the segment. This assumption is reasonable when the segment is short and maintains similar features, including both traffic volume and roadway geometric characteristics. This approach has been proved to be a reliable method by comparing the travel time estimates with empirical observations (Zhao 2011). In their research, the segment being studied was divided into several shorter sub-segments. The travel time of each sub-segment was obtained by dividing the sub-segment distance by the average truck speed along the sub-segment. Travel time of the entire link was the sum of the travel time of each sub-segment. The result was compared with both empirical GPS observations and estimates based on loop detector data. It is found that the approach is sufficiently accurate to estimate truck travel time on freeways.

3.4 Evaluate results

Mean absolute percentage error (MAPE), which is widely used as a measure to quantify the difference between the estimated value and the observed value, is chosen to evaluate the accuracy of the prediction, as shown in equation 8. In this study, the observed travel time is defined as the estimates obtained by dividing segment distance by average truck speed from GPS data. The MAPE value of the proposed approach is compared with the MAPE values of the BRP function and the Akçelik function. A lower MAPE value represents more accurate prediction of truck travel time.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{TT_i - TT_i'}{TT_i} \right| \times 100\% \quad (8)$$

where n = total number of examples,

TT_i = observed travel time,

TT_i' = model predicted travel time.

4. Case Studies

4.1 Data preparation

Two traffic datasets from different locations in Washington State in the USA were collected to demonstrate the proposed approach: Interstate-5 (I-5) northbound between milepost (MP) 158 and 161 in the City of Seattle and Interstate-405 (I-405) northbound between MP 8 and 10 in the City of Bellevue. Figure 1 shows the locations of the two segments in Great Seattle Area. Both are high capacity roadways and critical connectors for the region and are similar in design and operations to many high capacity multi-lane highways in Europe. The I-5 segment being studied is a 5-lane (each direction) interstate highway served 195,000 vehicles daily (both directions) in 2013 (WSDOT, 2015a). The mean truck travel speed was 54.7 mph and there were 10% of trucks traveling below 60% of posted speed of 60 mph between September 2010 and September 2011 (WSDOT, 2015b). The I-405 segment is a 5-lane (each direction) interstate highway carried 150,000 vehicles daily (both directions) in 2013 (WSDOT, 2015a). The mean truck travel speed was 49.2 mph and there were 15% of trucks traveling below 60% of posted speed of 60 mph (WSDOT, 2015b). With substantial growth of travel demand in Great Seattle area, traffic along these two segments experiences considerable delay. In addition, the traffic is not stable during different times-of-day. It is unknown how truck travel time would be improved due to potential investment in infrastructure. These two interstate segments were selected as case studies to demonstrate how truck GPS data can be used to forecast truck travel time on a freeway associated with transportation investments.

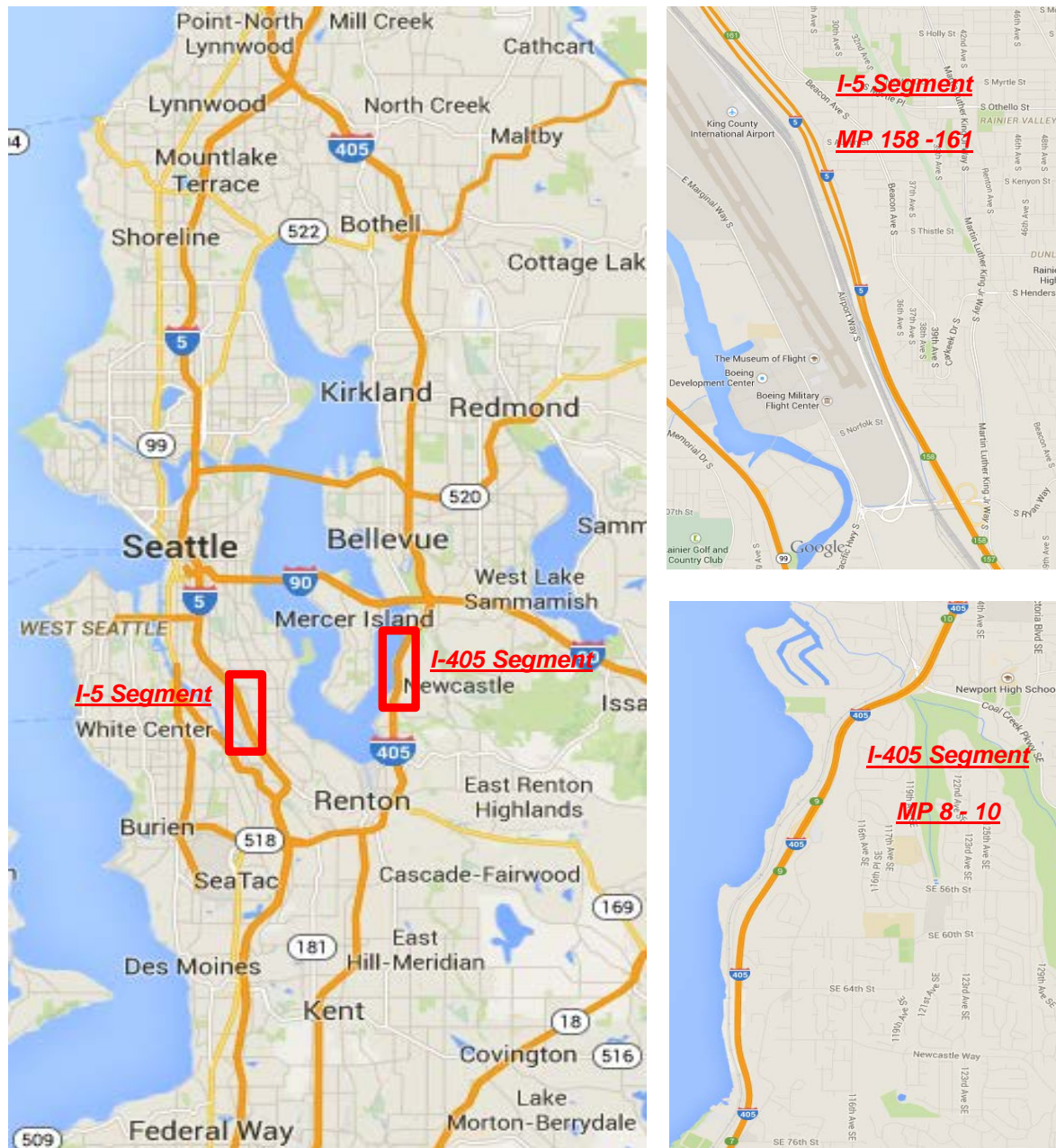


Figure 1. Locations of the two case studies

Truck Speed

Truck speed used in this research was retrieved from GPS devices installed in commercial vehicles traveling along the two selected segments. Data was collected anonymously from May 2012 to July 2012. The GPS data was reported every 2-15 minutes. Information provided by GPS includes a unique device ID, latitude and longitude, instantaneous truck speed, truck heading direction, and timestamp (time and date). Vehicle composition and truck trip purposes are unknown due to the anonymity of data collection. Data was cleaned and geocoded to the freeway network in the ArcGIS environment. More details of data processing can be found in McCormack (2011). GPS data was aggregated into 1 hour bins for each freeway segment to get average truck speed along the link.

Roadway Density

Roadway density was obtained by dividing traffic volume by truck speed. Traffic volume was collected by dual-loop detectors deployed in the right-most lane. The raw loop data provides

traffic counts every 20 seconds. Traffic count data was also aggregated into every 1 hour. Case study I contains six loop detectors deployed at MP 158.21, 158.92, 159.2, 159.96, 160.4 and 160.97. Case study II contains five loop detectors deployed at MP 8.03, 8.4, 8.9, 9.36 and 9.75. Traffic volume was estimated as the averaged value of loop detector collections along the segment.

4.2 Case study

4.2.1 Case study I

A 3-mile stretch of northbound I-5 in the City of Seattle, between MP 158 and MP 161 was selected as case study I. Both truck GPS data and loop data were collected between May 2012 and July 2012. The data set was divided into a training set (May 2012 and June 2012) and a testing set (July 2012). Truck speeds along the segment was retrieved from GPS data. Traffic volume was calculated as the averaged traffic volume recorded by the six dual loop detectors. Density was obtained by dividing traffic volume by truck speed. Figure 2 displays the truck speed-density plot of the training dataset.

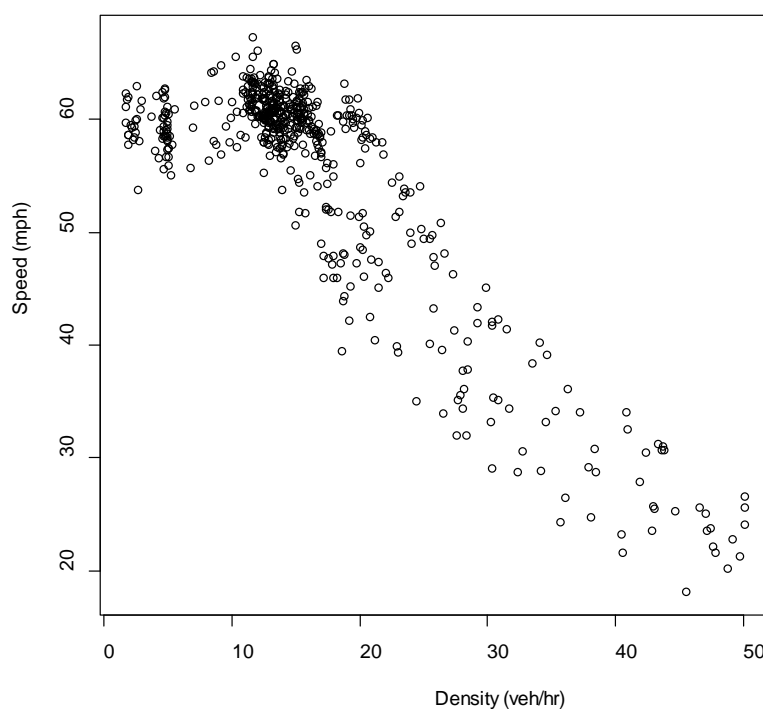


Figure 2. Case study I truck speed-density plot

As shown in Figure 2, trucks maintain a constant speed around 60 mph when segment density is less than 10 vehicles/mile and speed drops significantly while density increases, with the lowest observed speeds of approximately 20 mph. The K-means algorithm was employed to classify dataset into different clusters representing various traffic regimes. It is clear from Figure 2 that there are at least two traffic regimes, and may be more as the speed decreases at different rates with the increase of density. The appropriate number of clusters is often ambiguous, and depends on the distribution of observations in a dataset and the desired resolution of the user. Meanwhile, the number should not be too many for convenient use of the model. Thus the authors conducted the cluster analysis with two clusters and three clusters respectively, and compare the results in the following sections.

Two Clusters

Figure 3 and Table 1 present the clustering results when there are two clusters. The first cluster characterizes the free-flow traffic regime, in which trucks travel at around 60 mph when segment truck density is less than 10 vehicles/mile. The clustering result shows that the average truck

speed of cluster 1 is equal to 60 mph, and average density is 4.87 vehicles/mile. The second cluster represents the non-free flow condition where truck speed starts to decrease when density is greater than 10 vehicles/mile and drops continuously with the increase in segment density. The average truck speed and segment density of the second cluster are 54.63 mph and 18.46 vehicles/mile respectively. It should be noted that the cluster numbers here are only used to identify each specific cluster.

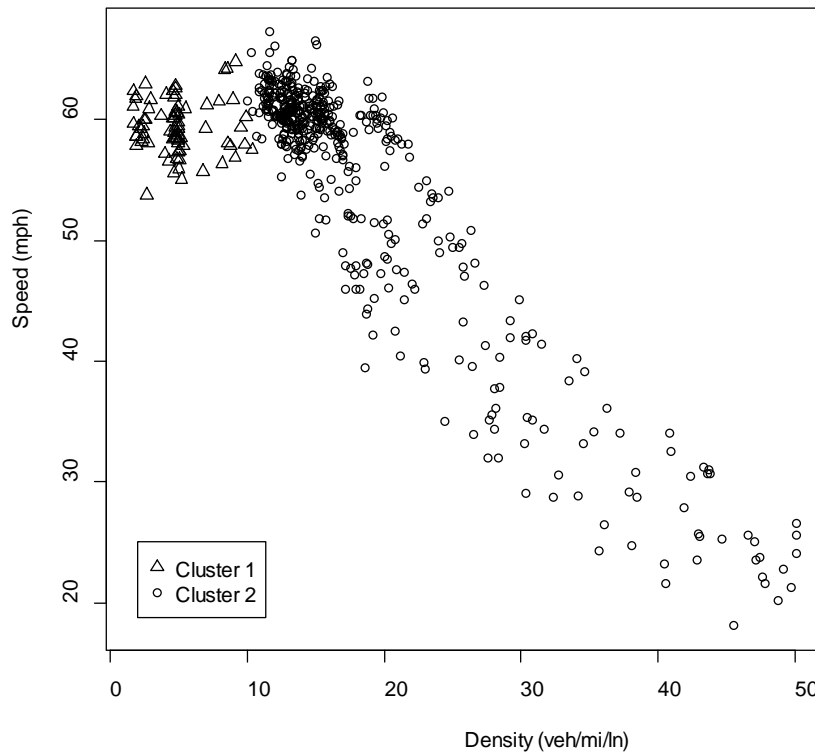


Figure 3. Case study I two clusters truck speed-density plot

Table 1. Case study I cluster centers of two clusters analysis

	Cluster 1	Cluster 2
Truck Speed (mph)	60	54.63
Density (vehicles/mile)	4.87	18.46

For cluster 1, trucks travel at the average of 60 mph regardless of the segment density. For cluster 2, truck speed is a dependent variable of density. The authors fitted the data using linear, logarithmic, and exponential models which were applied in the rational speed-flow relationships shown in Equation 5 to 7. It is found that the exponential function provides the best fit of the observed data with the greatest R-squared value, and the regression results are summarized in Table 2. All parameters are significant with P-values less than 0.0005. The truck speed-density relationship of the test dataset is given in Equation 9.

Table 2. Case study I the second cluster fitted results of two clusters analysis

	Coefficients	Standard Error	t Stat	P-value
Intercept	4.478	0.009	469.177	<0.0005
Density	-0.027	0.000	-58.459	<0.0005

$$\begin{cases} u = 60 & \text{for } k \leq 10 \\ u = \exp(4.478 - 0.027k) & \text{for } k > 10 \end{cases} \quad (9)$$

Data collected in July 2012 was used to evaluate the proposed approach. Hourly traffic volume was retrieved from loop detector data and averaged hourly truck speed was calculated from truck GPS data. Truck travel time obtained from dividing the segment distance by observed truck GPS speed was used as the ground truth travel time to evaluate the accuracy of the proposed approach. The authors also employed the BPR function and Akçelik function to estimate travel time, and compared with the ground truth travel time to calculate the corresponding MAPE values and evaluate the accuracy of each method. As shown in Table 3, the MAPE value of the proposed speed-density based approach is 6.16%, less than the MAPE values of BPR and Akçelik methods of 11.52% and 11.60% respectively. This result indicates that the proposed approach generates less deviation between travel time estimates and observations, and therefore performs better than the existing BPR method and Akçelik method.

Table 3. Case study I MAPE values of each travel time prediction method

	MAPE value
Speed-density method (two clusters)	6.16%
BPR method	11.52%
Akçelik method	11.60%

Three Clusters

Figure 4 and Table 4 show the clustering results with 3 clusters. Similar to the two clusters results, cluster 1 represents the free-flow traffic regime, in which traffic density is low and truck travel at about 60 mph when density is less than 11 vehicles/mile. The speed is constant and not affected by density. The average truck speed and density of cluster 1 is 60 mph and 4.94 vehicles/mile respectively. Truck speed in cluster 2 and cluster 3 decreases considerably with the increase of density. Cluster 2 features a high speed and intermediate density phase when density is between 11 and 25 vehicles/mile, and cluster 3 characterizes a low speed and high density congested phase when density is greater than 25 vehicles/mile. For cluster 2, the average speed and density are 58.77bmph and 15.26 vehicles/mile. For cluster 3, the average speed and density are 32.33 mph and 35.71 vehicles/mile.

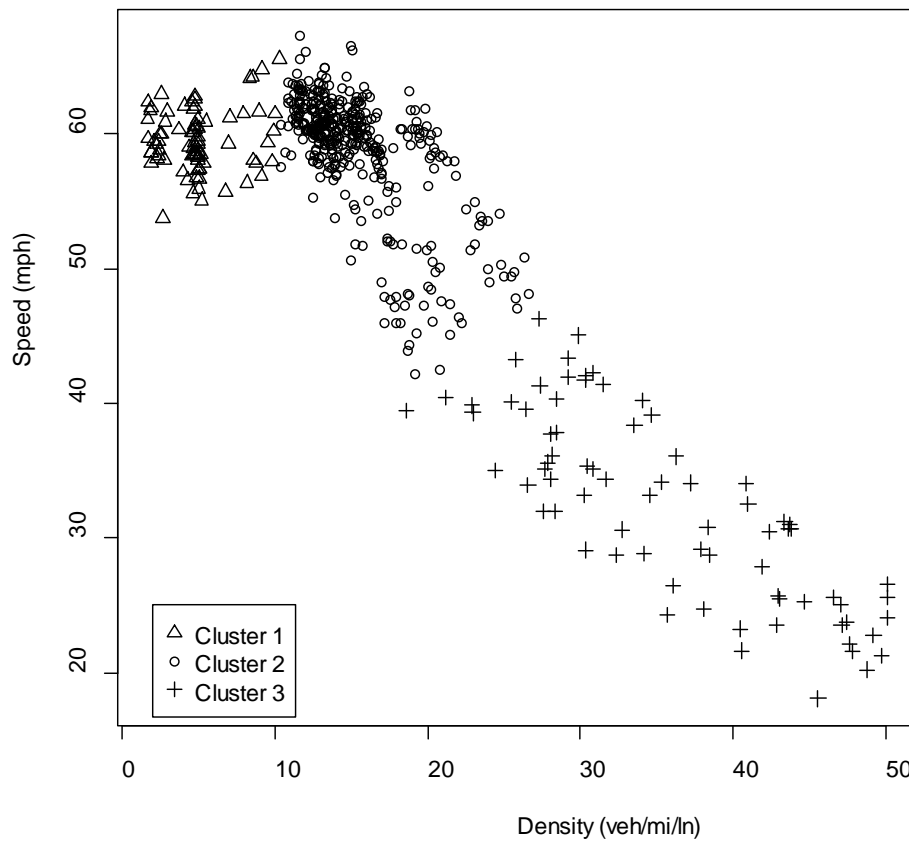


Figure 4. Case study I three clusters truck speed-density plot

Table 4. Case Study I Cluster Centers of Three Clusters Analysis

	Cluster 1	Cluster 2	Cluster 3
Truck Speed (mph)	60	58.77	32.33
Density (vehicles/mile)	4.94	15.26	35.71

As shown in Figure 4, the rate at which speed decreases differs between cluster 2 and 3. The linear, logarithmic and exponential models were tested to fit the cluster 2 and 3 data. It is found that the linear function fits cluster 2 data best and exponential function fits the cluster 3 data best, the fitting results are presented in Table 5. All parameters are statistically significant. Truck speed-density relationships are given in Equation 10.

Table 5. (a) Case study I second cluster fitted results and (b) third cluster fitted results

	Coefficients	Standard Error	t Stat	P-value
Intercept	72.709	0.794	91.569	<0.0005
Density	-0.914	0.051	17.975	<0.0005

(a)

	Coefficients	Standard Error	t Stat	P-value
Intercept	4.238	0.068	61.926	<0.0005
Density	-0.022	0.002	11.804	<0.0005

(b)

Truck speed and density relationship:

$$\begin{cases} u = 60 & \text{for } k \leq 10 \\ u = 72.709 - 0.914k & \text{for } 11 < k < 25 \\ u = \exp(4.238 - 0.022k) & \text{for } k \geq 25 \end{cases} \quad (10)$$

Similar to the previous analysis, the authors evaluated the proposed model using the test dataset and calculated the MAPE values. The MAPE value of the proposed approach is 5.55% as shown in Table 6. This value is less than the corresponding values of the BPR method and the Akçelik method, which are 11.52% and 11.60% respectively. This result reveals that the proposed approach generates, by a substantial margin, more accurate results than the other two methods.

Table 6. Case study I MAPE values of the selected travel time prediction methods

	MAPE value
Speed-density method (three clusters)	5.55%
BPR method	11.52%
Akçelik method	11.60%

By comparing Table 3 and Table 6, the two clusters and three clusters analysis results show that the MAPE value is improved from 6.61% to 5.55%. While the three clusters approach provides a slightly more accurate result, it also requires more data analysis efforts. While the user is entitled to choose the number of clusters appropriate for their study, for this case study, no significant improvement is observed when using three clusters instead of two clusters, and the case study is carried forward with the two clusters approach.

4.2.2 Case study II

Case study II is a 2-mile segment of I-405 northbound between MP 8 and 10. Traffic volume was the averaged value of data collected by the five loop detectors deployed along the rightmost lane. The speed-density plot is displayed in Figure 5. Similar to case study I, trucks travel at a constant speed in free-flow traffic pattern. Truck speed decreases when density is greater than 20 vehicles/mile. Both two clusters and three clusters analyses were performed to identify the appropriate number of clusters for this dataset.

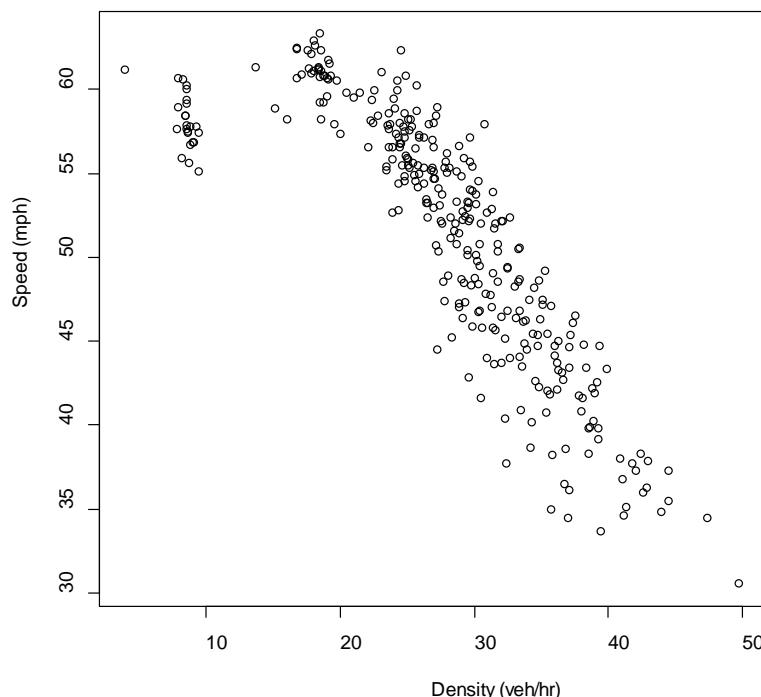


Figure 5. Case study II truck speed-density plot

Two Clusters

Figure 6 and Table 7 present the clustering results with two identified clusters. Cluster 1 features free-flow phase, in which trucks travel at a constant speed. According to the cluster analysis result, the average speed and density of cluster 1 is 58 mph and 9.12 vehicles/mile. For cluster 2,

truck speed starts to decline when density is greater than 16 vehicles/mile. The average speed and density are 50.88 mph and 29.39 vehicles/mile respectively.

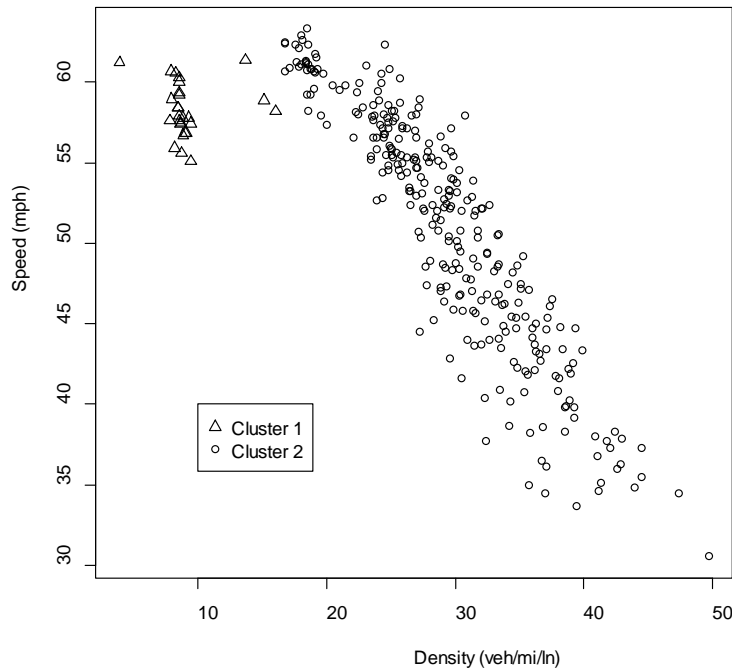


Figure 6. Case study II two clusters truck speed-density plot

Table 7. Case study II cluster centers of two clusters analysis

	Cluster 1	Cluster 2
Truck Speed (mph)	58	50.88
Density (vehicles/mile)	9.12	29.39

To fit the data of cluster 2, the linear, logarithmic and exponential models were tested, and the adjusted R-squared values of each model indicate that the linear model provides the best fit. The model results are presented in Table 8. The truck speed-density relationship is given in Equation 11.

Table 8. Case study II the second cluster fitted results of two clusters analysis

	Coefficients	Standard Error	t Stat	P-value
Intercept	81.88	0.82	99.70	<0.0005
Density	-1.05	0.03	-38.65	<0.0005

$$\begin{cases} u = 58 & \text{for } k \leq 16 \\ u = 81.88 - 1.05k & \text{for } k > 16 \end{cases} \quad (11)$$

The MAPE values of the proposed approach, BPR function and Akçelik function were calculated and the results are summarized in Table 9. The proposed approach generates the least MAPE value, and therefore performs better than the other two approaches.

Table 9. Case study II MAPE values of the selected travel time prediction methods

	MAPE value
Speed-density method (two clusters)	7.33%
BPR function	13.75%
Akçelik method	14.60%

Three Clusters

The three clusters were tested, and the speed-flow plot is displayed in Figure 7 and Table 10. While cluster 1 characterizes the free-flow traffic regime, cluster 2 represents the intermediate phase and cluster 3 features the congested phase. The authors tested linear, logarithmic and exponential models and find that none of them is able to delineate the dataset of cluster 2 and 3 well given the R-squared values are all less than 0.5. Thus the authors concluded that two clusters analysis is better than three clusters for this specific data of case study II.

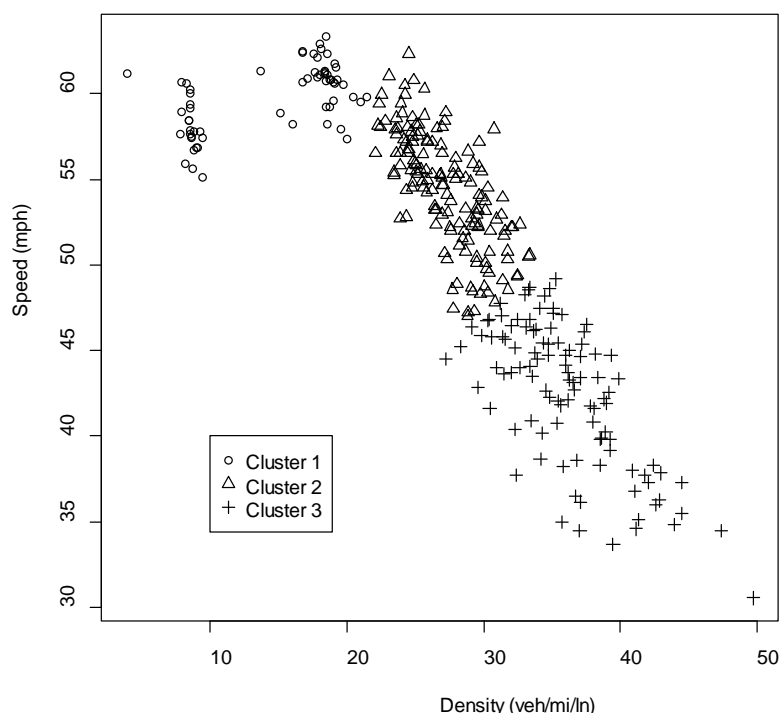


Figure 7. Case study II three clusters speed-density plot

Table 10. Case study II cluster centers of three clusters analysis

	Cluster 1	Cluster 2	Cluster 3
Truck Speed	59.73	54.35	42.46
Density	14.49	27.31	36.06

The above two case studies illustrate how the proposed multi-regime speed-density based approach can be used to forecast truck travel time. The analysis results indicate that the proposed approach is superior to the traditional BPR method and Akçelik method, and is able to forecast more accurate travel time. The number of clusters can be determined by both the distribution of data and the desired resolution of the user. The increase of number of clusters is able to improve the travel time prediction accuracy, but will involve additional data processing efforts and model application complexity. For both case studies, two clusters are able to provide substantial improvements over current methods used to predict truck travel time.

Despite the fact that both case studies considered freeways in the Puget Sound, the speed-density relationships shown in Equations 9 and 11 are different. Figure 8 shows the speed-density plot of the two datasets. It is noted that two case studies have distinct speed-density relationships. For case study I, speed starts to decline when density is greater than 10 vehicles/mile, while the breakpoint of case study II is around 16 vehicles/mile. Further, when density exceeds the breakpoint, dataset 1 has a convex shape and an exponential model provides the best fit, while dataset 2 displays a straight and linear relationship. The deviation of the speed-density distributions is associated with several characteristics of each segment, including roadway geometric features and travel demand distribution. Thus, the objective of this paper is not to

develop a generalized model that is applicable for any location as it is less accurate to select one model to fit different datasets. Given this, and the simplicity of the approach, we recommend users to apply the clustering and best-fit modelling approach, and develop their own equations for different locations.

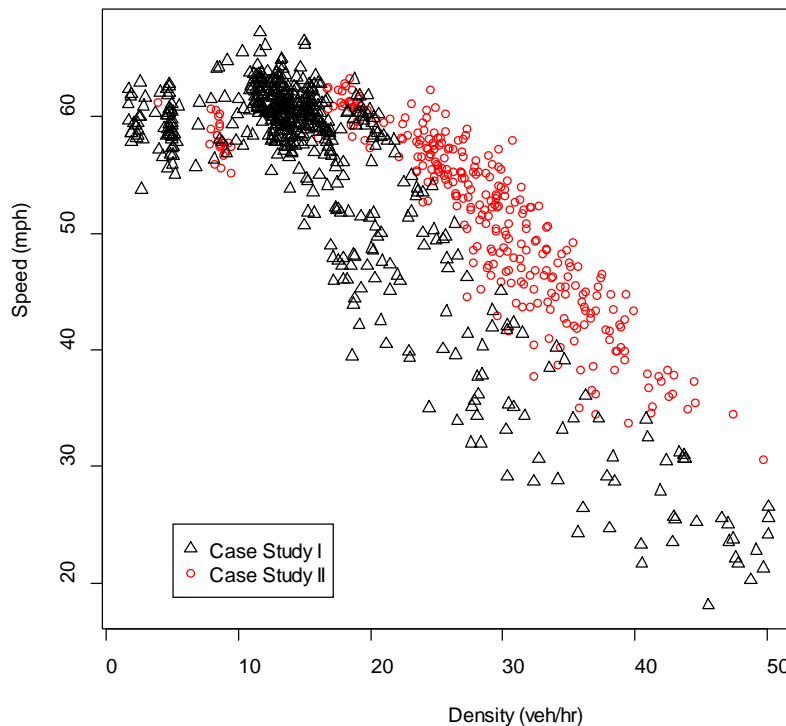


Figure 8. Speed-density plot of case study I and case study II

5. Conclusions

This paper proposes a multi-regime speed-density relationship based approach to predict freeway truck travel time using empirical truck probe GPS data and loop detector data. The impacts of both truck and passenger vehicle densities are included as the density data used in this study is mixed traffic density collected by loop detectors. The K-means cluster analysis algorithm was employed to determine the breakpoints of different traffic regimes. Each cluster was fitted using linear, logarithmic, and exponential models, and the model with the highest R-squared value was selected. The parameters of the best models for both cases are all statistically significant. The travel time estimates were compared with estimates calculated based on empirical truck GPS speed data, and the mean absolute percentage error was calculated. This was compared with the widely used BPR model and the Akçelik model. It is found that the new approach is able to estimate more accurate travel times than traditional methods given it generates the least MAPE values for both case studies. The truck GPS data employed in this study was not collected specially for predicting travel time, but for trucking company fleet management. With the growing market penetration of GPS technology in the trucking industry in both North American and Europe, this paper is a great asset to demonstrate how this type of GPS data can be used to produce better travel time estimates. In addition, it should be noted that this paper does not intend to develop a generalized result that fits different North American or European datasets since travel time is determined by many other factors in addition to traffic density. We recommend users to apply the proposed approach to develop their own equations for different locations.

The authors investigated the appropriate number of clusters when segment the data using the K-means algorithm. For case study I, the two-cluster identification is recommended since it is easier

to use and still provides reasonably high accuracy of estimates. For case study II, the two-cluster identification is recommended as well since the commonly applied speed-density formats do not fit the three-cluster clustering results. The analysis reveals that the number of clusters is determined by the distribution of data and the resolution desired by users. In the case studies evaluated, the more clusters are classified, the less deviation is obtained. The two clusters analysis is recommended when the improvement from three to two clusters is subtle.

The predicted travel time is usable to support freight planning and project prioritization effort as conducted in the USA and Europe. For instance, the speed-density relationships helps highway managers understand how truck speeds changes in response to different traffic regimes, and therefore managers can adjust traffic volume by imposing tolls or changing existing toll rates to manage the system to reach the desired speed. Another example of application is to forecast truck travel time associated with traffic density changes resulting from different freight investments and prioritize projects based on their impacts. More specifically, one can apply this approach to generate the multi-regime speed-density relationships based on GPS and loop data, and estimate the corresponding post-project traffic density. Although the GPS data may be provided by different vendors than that described in this paper, the data formats are consistent with those identified in this paper.

References

- Akçelik, R. (1991). Travel Time Functions for Transport Planning Purposes: Davidson's Function, Its Time-Dependent Form and An Alternative Travel Time Function. *Australian Road Research* 21 (3), 49–59. (Minor revisions: December 2000)
- American Transportation Research Institute (ATRI) and Federal Highway Administration (FHWA). (2005). *Measuring travel time in freight-significant corridors. Rep. FHWA-HOP-05-036*, Washington, DC.
- Atlanta Regional Commission (2011). *The Travel Forecasting Model Set For the Atlanta Region 2010 Documentation*, http://www.atlantaregional.com/File%20Library/Transportation/Travel%20Demand%20Model/tp_arcmodeldocumentation_022212.pdf, accessed January 2014.
- Battelle (2011) FAF3 Freight Traffic Analysis, http://faf.ornl.gov/fafweb/Data/Freight_Traffic_Analysis/faf_fta.pdf, accessed January 2014
- Berg Insight (2014) Fleet Management in Europe – 9th Edition, M2M Research Series.
- Cambridge Systematics, Inc. and Jack Faucett Associates (2001). National Cooperative Highway Research Program 298, Truck Trip Generation Data, Washington, DC.
- Cambridge Systematics, Inc. (2007). *PSRC Travel Model Documentation (for version 1.0)*, http://www.psrc.org/assets/1511/model_doc_final_.pdf, accessed January 2014.
- Chien, S. I. J., & Kuchipudi, C. M. (2003). Dynamic travel time prediction with real-time and historic data. *Journal of transportation engineering*, 129(6), 608-616.
- D'Angelo, M. P., Al-Deek, H. M., & Wang, M. C. (1999). Travel-time prediction for freeway corridors. *Transportation Research Record: Journal of the Transportation Research Board*, 1676(1), 184-191.
- Dowling Associates, Inc (2000) Washington State Department of Transportation Mobility Project Prioritization Process Benefit-Cost Software Users' Guide. http://www.wsdot.wa.gov/mapsdata/travel/pdf/Mobility_Users_Guide_2001.PDF, accessed January 2014.
- Dowling, R. G., Singh, R., & Cheng, W. W. K. (1998). Accuracy and Performance of Improved Speed-Flow Curves. *Transportation Research Record: Journal of the Transportation Research Board*, 1646(1), 9-17.

Drake, J. S., J. L. Schofer, and A. D. May (1967). A Statistical Analysis of Speed Density Hypotheses. Proc., Third International Symposium on the Theory of Traffic Flow. Elsevier North-Holland, Inc., New York

Edie, L. C. (1961) Car Following and Steady-State Theory for Non-Congested Traffic. *Operations Research*, Vol. 9, 66-76.

Federal Highway Administration (FHWA) (2013). National Performance Management Research Data Set, http://www.ops.fhwa.dot.gov/freight/freight_analysis/perform_meas/vpds/npmrdsfaqs.htm, accessed January 2014.

Figliozi, M. A., Wheeler, N., Albright, E., Walker, L., Sarkar, S., & Rice, D. (2011). Algorithms for Studying the Impact of Travel Time Reliability Along Multisegment Trucking Freight Corridors. *Transportation Research Record: Journal of the Transportation Research Board*, 2224(1), 26-34.

Greaves, S. P. and Figliozi, M. A. (2008). Collecting Commercial Vehicle Tour Data with Passive Global Positioning System Technology: Issues and Potential Applications. *Transportation Research Record: Journal of the Transportation Research Board*, 2049(1), 158-166.

Greenshields, B. D. (1935), A Study in Highway Capacity, *Highway Research Board Proceedings*, 14, 448-477.

Greenberg, H. (1959). An analysis of traffic flow, *Operation Research*, 7, 79-85.

Hartigan, J.A. (1975). Clustering algorithms. John Wiley & Sons, Inc.

Kerner, B. S., & Rehborn, H. (1996). Experimental properties of complexity in traffic flow. *Physical Review Series*, 53, R4275-R4278.

Liao, C.F. (2014) *Using Truck GPS Data for Freight Performance Analysis in the Twin Cities Metro Area*. <http://www.dot.state.mn.us/research/TS/2014/201414.pdf>. Accessed Jan 2014.

Lin, H. E., Zito, R., & Taylor, M. (2005). A review of travel-time prediction in transport and logistics. In *Proceedings of the Eastern Asia Society for transportation studies*, Vol. 5, 1433-1448.

Ma, X., McCormack, E. D., & Wang, Y. (2011). Processing Commercial Global Positioning System Data to Develop a Web-Based Truck Performance Measures Program. *Transportation Research Record: Journal of the Transportation Research Board*, 2246(1), 92-100.

McCormack, E., and Hallenback, M. E. (2006). ITS Devices Used to Collect Truck Data for Performance Benchmarks. *Transportation Research Record: Journal of the Transportation Research Board*, No 1957, 43-50.

McCormack, E., Scharnhorst, E., and Zhao, W. (2011). Using GPS truck data to identify and rank bottlenecks in Washington State. *Rep. prepared for Washington State Department of Transportation*, Olympia, WA.

McCormack, E. and Aakre, A. (2014). Using ITS devices to Analyze Norwegian Truck Travel, paper presented at the *10th ITS European Congress*, June 2014, Helsinki, Finland.

McFarland, W. F., J. Memmott, and M. Chui (1993). *Microcomputer Evaluation of Highway User Benefits - Final Report for NCHRP 7-12*. Texas Transportation Institute, Texas A&M University. College Station, TX.

Pluvinet, P., Gonzalez-Feliu, J., & Ambrosini, C. (2012). GPS data analysis for understanding urban goods movement. *Procedia-Social and Behavioral Sciences*, 39, 450-462.

R Software Package "Cluster", <http://cran.r-project.org/web/packages/cluster/cluster.pdf>, accessed March 2014.

Sun, L. and Zhou, J. (2005). 'Development of Multiregime Speed-Density Relationship by Cluster Analysis'. *Transportation Research Record*, No. 1934, 64-71.

Transportation Research Board (2000). Highway Capacity Manual (2000). <http://www.trb.org/Main/Blurbs/164718.aspx>, accessed Jan 2014.

Underwood, R. T. (1961). Speed, Volume, and Density Relationships, *Quality and Theory of Traffic Flow*, New Haven, Connecticut, 141-188.

U.S. Bureau of Public Roads (1964). *Traffic Assignment Manual*, Washington, DC.

Van Lint, J. W. C., Hoogendoorn, S. P., & van Zuylen, H. J. (2005). Accurate freeway travel time prediction with state-space neural networks under missing data. *Transportation Research Part C: Emerging Technologies*, 13(5), 347-369.

Wang, H.Z., Li, J., Chen, Q.Y., and NI, D.H. (2009). Speed-Density Relationship: From Deterministic to Stochastic, Transportation Research Board 88th Annual Meeting <http://people.umass.edu/ndh/Publications/C21.pdf>

Wang, Z. and Goodchild, A.V. (2014). GPS Data Analysis of the Impact of Tolling on Truck Speed and Routing: Case Study in Seattle, Washington. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2411, Transportation Research Board of the National Academies, Washington, D.C., 2014, 112-119.

Wang, Z., Goodchild, A.V. and McCormack, E. (2015). Measuring Truck Travel Time Reliability Using Truck Probe GPS Data, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 2015, 1-10.

Washington State Department of Transportation (WSDOT) (2015a). WSDOT Traffic Volume Map Application, <http://www.wsdot.wa.gov/mapsdata/tools/trafficplanningtrends.htm>

Washington State Department of Transportation (WSDOT) (2015b). Washington State Freight Map, <http://www.wsdot.wa.gov/data/tools/geoportal/?config=freight>

Yeon, J., Elefteriadou, L., & Lawphongpanich, S. (2008). Travel time estimation on a freeway using Discrete Time Markov Chains. *Transportation Research Part B: Methodological*, 42(4), 325-338.

Zhao, W., Goodchild, A. V., & McCormack, E. D. (2011). Evaluating the accuracy of spot speed data from global positioning systems for estimating truck travel speed. *Transportation Research Record: Journal of the Transportation Research Board*, 2246(1), 101-110.